

## Article

# What role for human capital in the growth process: new evidence from endogenous latent factor panel quantile regressions

Kostov, Phillip and Le Gallo, Julie

Available at <https://clok.uclan.ac.uk/23984/>

*Kostov, Phillip orcid iconORCID: 0000-0002-4899-3908 and Le Gallo, Julie (2018) What role for human capital in the growth process: new evidence from endogenous latent factor panel quantile regressions. Scottish Journal Of Political Economy, 65 (5). pp. 501-527. ISSN 0036-9292*

It is advisable to refer to the publisher's version if you intend to cite from the work.  
<http://dx.doi.org/10.1111/sjpe.12196>

For more information about UCLan's research in this area go to  
<http://www.uclan.ac.uk/researchgroups/> and search for <name of research Group>.

For information about Research generally at UCLan please go to  
<http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the [policies](#) page.

# **What role for human capital in the growth process: new evidence from endogenous latent factor panel quantile regressions**

Philip Kostov<sup>a</sup> and Julie Le Gallo<sup>b</sup>

<sup>a</sup> Lancashire School of Business and Enterprise, Greenbank Building, GR032, Preston, PR1 2HE, UK.

<sup>b</sup> [CESAER](#), [Agrosup Dijon](#), DSHS Longelles, 26 Boulevard Petitjean, 21079 Dijon Cedex, France

## **Abstract:**

The estimates for the human capital effect in cross-country growth regressions have been subject of considerable controversy. We argue that human capital is intrinsically a multidimensional construct. We construct human capital measure by combining available alternative proxies *via* confirmatory factor analysis. Using panel data endogenous quantile regression methods we analyse the whole conditional growth distribution by simultaneously accounting for the potential endogeneity of human capital and country specific effects. Our results conform to theoretical expectations and we are able to demonstrate the beneficial effect of both the measurement approach and the endogeneity correction on the derivation of theoretically consistent estimates.

**Key words:** human capital; cross-country convergence; latent factor; quantile regression; panel data

**JEL codes:** C21, C23, O47

## 1. Introduction

Cross-country growth regressions routinely employ educational variables as proxies for human capital. However, the significance of education measures in growth regressions has been source of considerable controversy. While the theoretical arguments of why it should be important in driving economic growth are compelling (Mankiw et al., 1992), the empirical evidence has been mixed (Krueger and Lindhal, 2001; Pritchett, 2001). Indeed, there are several potential problems with used educational variables.

The first one is the imperfect nature of such measures. There has been considerable debate on the appropriateness of different proxies for human capital and the quality of the available data. Different researchers have put forward new improved databases of such proxies. Examples include levels of educational attainment, such as the fraction of working age population in secondary school (see Nehru et al. 1995; de la Fuente and Doménech, 2006; Cohen and Soto, 2007; Lutz et al., 2007; Barro and Lee, 2010). However, these measures present a number of drawbacks (Folloni and Vittadini, 2010), in particular, not only the quantity but also the quality of years of education have an impact on the cognitive skills acquired and ultimately on growth (Wössman, 2003). Hence, these proxies should be designed to measure an intrinsically unobservable variable: the quality of human capital. While it is conceptually inconceivable that a single proxy would be able to successfully capture the quality of human capital, empirical research in cross-country growth regressions has nevertheless traditionally applied a single proxy for it. On the other hand, using several human capital variables can lead to issues with multicollinearity: it would be then tricky to disentangle the effect of several such measures used in the same model. Owing to the imperfect nature of such measures, it is therefore unsurprising that such models have been known to exhibit considerable heterogeneity and nonlinearities within the growth process (see Temple, 2001; Kalaitzidakis et al. 2001; Sianesi and Van Reenen, 2003; Sunde and Vischer, 2015).

The conceptual model underlying the cross-country growth regression suggests an unobservable human capital variable that could be potentially measured by a number of different proxies. One possibility to overcome this difficulty is to consider human capital as a factor variable transforming part of the design matrix of the growth regression to incorporate a confirmatory factor analysis structure. Hence, this paper approaches this problem by using a general structural equation model (SEM) framework. SEMs have been

used in growth regressions to estimate the impact of variables, for which only imperfect proxies exist such as policy variables (Brumm, 1997), well-being variables (Cracolici et al., 2010) and various factors such as creative capital, entrepreneurship or leadership (Aroca et al., 2014). However, to the best of our knowledge, these models have not been used to deal with the specific issue of human capital. In this paper, based on the CANA database (Castellacci and Natera, 2011), we use the main educational variable 'Mean years of schooling' together with 'Public Expenditure on Education', and 'Primary teacher-pupil-ratio' to construct a confirmatory factor analysis of the human capital factor. Our approach can therefore be seen as a macroeconomic extension of the approach set out by Dagum and Slottje (2000) who, at a microeconomic level, consider human capital as a multidimensional non-observable construction of personal ability, home and social environments and investments in education of the household head and spouse.

The second issue with educational proxies and the human capital factor that is to be derived from them is their potential endogeneity. We propose to address this endogeneity *via* instruments. Search for instruments for human capital in itself is a problematic area as issues of validity of instruments and identification are difficult to deal with (Temple, 1999). Here, we suggest using lagged enrolment rates to address the endogeneity problem. Enrolment rates can be viewed as an aggregate determinant of educational achievement and therefore should satisfy the exogeneity requirement. However, they only affect it after a considerable lag of time and therefore should be excluded from defining human capital directly. Hence, lagged enrolment rates appear to satisfy the exogeneity and the exclusion assumptions necessary for valid instruments. One potential issue with enrolment rates is that they are known to be changing very slowly which can lead to weak instruments. Therefore, we suggest ways to test for weak instrumentation in the empirical model.

Additionally, we explicitly tackle the general issue of parameter heterogeneity in growth regressions. Indeed, considerable evidence for parameter heterogeneity has been found in cross-countries studies (Durlauf et al., 2009). Applying a quantile approach to convergence analysis makes it possible to use each estimated quantile to describe a particular segment of the conditional distribution of income growth (Cunningham, 2003; Barreto and Hughes, 2004); Canarella and Pollard, 2004; Foster, 2008; Ram, 2008;

Dufrenot et al., 2010). The quantile approach is therefore of particular interest since the conditional convergence hypothesis is defined in terms of conditional growth and (conditional) quantile regression explicitly models conditional growth.

It is then possible to provide a more complete description of the relationship between income growth rate and initial level of per capita income, human capital and other variables. Moreover, note that while the above studies only apply cross-sectional version of quantile regressions, we fit panel quantile regressions that allow controlling for unobserved country-specific effects.

This paper hence adds to the literature by proposing a general panel quantile model with one endogenous variable: human capital. The results support the theoretical conjecture with regard to both the convergence process and the effect of human capital. Indeed, we find evidence for conditional convergence at all quantiles of the conditional growth distribution, with coefficients for the lagged income significant and negative varying between -0.05 and -0.01 depending on the quantiles and the specifications. Human capital has a positive and significant effect, particularly for the conditionally slower growing countries. Its effect varies greatly, depending whether it is considered exogenous or corrected for endogeneity. In particular, we demonstrate how ignoring endogeneity and using alternative (standard) measures for human capital leads to reduced support for the theoretical model.

This paper proceeds as follows. In the next section, we discuss the challenges in measuring human capital. Then, we elaborate the model and estimation issues (Section 3) followed by a presentation of the data used in the empirical analysis (Section 4). Section 5 outlines our estimation results together with comparisons with some alternative models. Finally, Section 6 concludes.

## **2. Human Capital conceptualisation**

From a theoretical point of view, human capital, in particular the part obtained through education, is a critical determinant of economic growth. Indeed, more skilled and productive workers increase an economy's output of goods and services and facilitate the absorption of advanced technologies. Hence, it comes as no surprise that considerable attention has been given to the most relevant ways to measure this particular input in

order to make cross-country comparisons (Wössman, 2003; Le et al., 2003; Folloni and Vittadini, 2010). Because human capital is intangible, and its stock is not directly observable, all estimates of this stock must be constructed indirectly.

With respect to growth, human capital is implicitly or explicitly defined in relation to productivity. Indeed the United Nations (1953) defined investment in human capital as investments made to increase the productivity of the labour factor. Hence, the idea underlying the concept of human capital is that economic output (GDP) can be increased not only by the conventional capital stocks, but also through increasing the quality of human capital. The latter implicitly means the productive capacity of humans which can be expressed (i.e. measured) in terms of characteristics such as education, work experience, acquisition of knowledge, health standards as well as many other intangible factors that affect labour productivity. The OECD report (1998) follows the same line by defining human capital as the 'knowledge, skill, competencies and attributes embodied in individuals that are relevant to economic activity'. Therefore, measuring human capital means measuring labour productivity, or alternatively the ability of a person to produce labour income. This leads to a dual possible approach to measuring human capital. It involves both costs (investment in human capital) such as education, health and quality of life improvements and labour earnings capacity (future wages).

First, human capital can be measured by a costs (investment) approach. Typically, this involves measures such as spending on education. This approach can be traced back to the cost-of-production method of Engel (1883) who estimated human capital as the child rearing costs to their parents (Le et al. 2003). The underlying assumption to such an approach is that the value of the human capital embodied in a person is equal to the cost of producing that same wealth. Examples can be found in Kendrick (1976) and Eisner (1985) for the United States but many criticisms have been put forward (Folloni and Vittadini, 2010) such as the absence of relation between the cost of production and quality of output and the fact that not all the components involved in the production of skills can be identified.

Second, human capital can be defined by its earning capacity. To this end, Mincer (1970, 1974), considering investment in human capital as a rational choice, specified an earnings function by employing a number of mathematically convenient but controversial counterfactual assumption. In simple terms, the Mincerian framework does not measure

human capital *per se* but defines it as a wage function in terms of determinants such as the length of school training and the total cost of post-school investment in training, health, and mobility. In essence, the pricing equation is a hedonic wage model, which defines earning capacity with regard to individual characteristics and includes the rate of return to schooling defining the optimality of human capital investments (with regard to the interest rate). Although intrinsically built on microeconomic logic, the Mincerian framework has been applied to macroeconomic data where the return to schooling has been viewed as a measure of the effect of human capital. It is clear that in such a framework the above effect would (in addition to the questionable assumptions embodied in the overall approach) crucially depend on the correct specification of the hedonic relationship.

The above illustrates that human capital is a complex, multifaceted concept. This creates some serious problems when trying to measure it directly or to evaluate its impact. We argue that trying to reduce it to a single attribute or variable is nothing more than wishful thinking. Yet, most empirical studies do exactly that and use some measure of the education level/achievement as a proxy for human capital. In cross-country growth studies the most popular measures are 'levels of educational attainment' and 'average years of schooling'. Such a reliance on education level is due, in part, to the difficulties of measuring the components of human capital and the unavailability of data, but also to the clear link to the Mincerian framework in which the effects of such a variable can be interpreted as return to schooling. Note however that as the above discussion shows, even within the assumptions of the Mincerian framework, such a variable clearly omits large part of the human capital. This point had already been noted by Temple (1999a, p. 139) who stated that '[t]he literature uses somewhat dubious proxies for aggregate human capital'. In particular, as pointed by Wössman (2003), specifying human capital by average years of schooling implicitly gives the same weight to any year of schooling acquired by an individual and the same weight to a year of schooling in any schooling system at any time, whereas there might be considerable differences in the quality of the education system, leading to major variations in the cognitive skills learned by the individuals during the years. These differences in the quality of the education system might have various origins. For instance, taking the McKinsey report on international achievement concluding that "the quality of an educational system cannot exceed the quality of its teachers", Hanushek et al. (2018) construct country-level measures of

teacher cognitive skills for 31 countries and find that student performance is strongly related to the substantial difference existing in teacher cognitive skills across countries. Hence, the expansion of education should also be accompanied by an increase in the cognitive skills, the knowledge capital of countries (Hanushek, 2016). Other studies tend to show that the teachers' salary levels are positively related to the cross-country differences in student outcomes (Woessman, 2005; Dolton and Marcenaro-Gutierrez, 2011).

However, as we have also shown above human capital should be viewed as a complex, multifaceted phenomenon various dimensions that are not directly observed. From a statistical point of view, it is hence a latent (unobservable) variable that can be related to a set of characteristics. Therefore, instead of looking for a single proxy for human capital, one can combine different characteristics (measures, determinants) to construct a new more reliable measurement for it. At a micro-economic level, Dagum and Slottje (2000) have defined human capital as a non-observable variable generated by personal ability, social environments, investments in education etc. and assumed that the effects of this multidimensional construct can be indirectly measured by the present value of the future earnings. This approach has several appealing properties. First, it is conceptually appealing since it avoids the reductionist logic of a single proxy practice and conforms to the implicitly multidimensional nature of the human capital definition. Second, it neatly combines the investment and earnings capacity views on human capital measurement. Instead of adopting one or the other, they can be combined in allowing one to extract as much useful information as possible, based on a set of deficient measured based on either of the investment and earning capacity points of view. Furthermore, it naturally leads to a structural equations model type of approach that has been used to model latent variables. Dagum and Slottje (2000) apply their method to estimate the 1983 human capital in the USA (see also Dagum et al. 2007 and Vittadini and Lovaglio, 2007 for discussions). Other applications involve Földvári and van Leeuwen (2006) for Eastern Europe and Le (2006) for New Zealand. Hereafter, we follow this approach at a macroeconomic level to investigate the impact of such defined human capital on growth and convergence.



### 3. Methodology

We base our analysis on a factor augmented endogenous panel quantile regression. This model assumes that the design matrix for the explanatory variables is a mixture of standard covariates and latent factors, the latter being modelled *via* confirmatory factor analysis (Harrington 2008; Brown, 2015). We choose to recover the factor structure from confirmatory factor analysis, rather than taking the agnostic approach that consists of using exploratory factor analysis, as in e.g. Ando and Tsay (2011). Indeed, in our case, we construct such an unobservable factor with the aim of measuring human capital. Since we use a number of proxy variables capturing different aspects of human capital, we have a prior knowledge of both the number of factors as well as their structure so that a confirmatory factor analysis is more relevant. The factor model is presented in section 3.1. Additionally, we allow for country individual effects and endogeneity. The associated estimation and inference issues in the context of our panel quantile model are presented in respectively sections 3.2 and 3.3.

#### 3.1. Models and assumptions

Formally, let  $t = 1 \dots T$  denote the index for time periods and  $i = 1 \dots N$  denote the index for cross-sectional observations. The cross-sectional observations are stacked on top of each other by year. Note by  $y$  the  $(NT, 1)$  vector containing the observations for the response variable and  $X$  and  $w$  are respectively the  $(NT, K)$  observations for the exogenous covariates and the  $(NT, 1)$  endogenous latent factor.<sup>1</sup> A confirmatory factor model is used to recover the unobserved factor  $w$  as in:

$$z = w\Lambda + u \quad (1)$$

where  $z$  is a stacked  $(NT, k)$  matrix of  $k$  measurement variables in each column;  $w$  is the latent factor  $(NT, 1)$  being measured;  $\Lambda$  is a  $(1, k)$  matrix of factor loadings and  $u$  is a  $(NT, k)$  matrix of *iid* normally distributed disturbances ordered by column. In (1) above, we use small letter for the unobserved factor to denote that it is a vector (since we only consider a single factor in this paper). Hence, in this particular case, we only have 1 latent factor with  $k$  proxy (indicator) variables for it. Then the matrix of the factor loadings  $\Lambda$

---

<sup>1</sup> Our presentation is for a model with a single such latent factor, but it is straightforward to extend it to more such factors.

is of dimensions  $(1,k)$ , *i.e.* the number of factors times the dimension of the proxy variables.

Using this factor model and, the structural quantile model, for quantile  $\tau$  can be written as:

$$y = \delta_\tau w + X\beta_\tau + \psi_\tau + \xi \quad (2)$$

with the same notations as before;  $\psi_\tau$  is a set of fixed country individual effects;  $\delta_\tau$  and  $\beta_\tau$  are the unknown parameters to be estimated. With respect to the error term, the only assumption we impose on  $\xi$  is the conditional linear quantile restriction:  $q_\tau(\xi_\tau | X) = 0$  for a given quantile  $0 < \tau < 1$ . Equation (2) is essentially the ‘fixed effects’ panel quantile regression model of Koenker (2004).

Finally, since the factor variable is endogenous, we can identify it by a set of exogenous instruments in  $S$  using the following reduced equation:

$$w = \mu_\tau + (X, S)\pi_\tau + \varepsilon \quad (3)$$

One can view the model in (1), (2) and (3) above as a combination of a standard endogenous linear quantile regression model for  $y$ , combined with a confirmatory factor analysis model for the latent factor  $w$ . We intentionally split the model specification in three separate equations in order to clarify its nature and the estimation approach we adopt. Equation (1) is a confirmatory factor analysis model, equation (2) is a panel quantile regression, while (3) identifies the variation in the endogenous variable *via* a set of instruments. If we only had (1) and (2) (*i.e.* if no endogeneity is present), this would reduce to a quantile regression model in which part of the design matrix is obtained *via* confirmatory factor analysis. Hence in this case, (1) corresponds to a measurement equation and (2) to the structural equation in a SEM type of model with the obvious difference that the structural equation describes conditional quantile (instead of conditional mean).

Burgette and Reiter (2012) have suggested a Bayesian approach to estimate such a combination for cross-sectional data, which estimates both these equations jointly. In our particular case, equation (2) is a panel data quantile regression and the same approach can be generalised by adopting appropriate priors for the country individual effects. It is also possible to further generalise the same approach by incorporating (3), *i.e.* estimating

jointly all three equations. Here, we follow a slightly different approach by estimating (1) separately and plugging in the estimated factor values in the rest of the model. The main reason for this choice is its computational advantages when dealing with endogeneity and with the estimation approach adopted for the panel data quantile model in equation (2). Since we estimate equation (1) as a preliminary step in the rest of the discussion we treat the endogenous factor  $w$  as given and focus on equations (2) and (3).

### *3.2. Individual effects in quantile models*

Equation (2) is a panel quantile regression. In our case, the individual effects  $\psi_i$  account for possible unobserved country heterogeneity. The standard approach to individual effects in linear panel model consists of applying some transformation (such as the within transformation, time differencing, orthogonal deviations etc.) that eliminates them allowing the transformed model to be estimated by standard methods. The quantile regression is however essentially a non-linear model. As a result, there is no transformation that can eliminate the individual effects, so that individual effects would need to be estimated directly. Such a strategy however can lead to a version of the incidental parameter problem (see e.g. Rosen 2009 for details). Hence, additional assumptions are needed to identify the individual effects in panel quantile regression models. Some of these identification strategies rely upon imposing restrictions (e.g. Canay, 2001) that compromise the generality of the quantile model.

The most popular (and extensively studied) approach to panel quantile regression is that of Koenker (2004) who suggested shrinking the 'fixed effects' via  $L1$  penalty to overcome the bias. The idea is very simple. While the introduction of individual effects increases the variability of their estimates, shrinking them towards a common value (via the  $L1$  penalty) helps reducing this variability. The asymptotics in this case relies on both  $T$  and  $N$  growing to infinity. In practice, implementing the 'fixed effect' quantile regression is relatively straightforward. The main stumbling block is the choice of optimal amount of shrinkage. Lamarche (2010) showed that under some regularity conditions, the regularised quantile estimator of Koenker (2004) is asymptotically unbiased. Then, choosing the amount of shrinkage that minimises the asymptotic variance is equivalent to minimising the average mean square error (AMSE) of the estimator. Nevertheless, selecting the optimal amount of shrinkage involves considerable computational costs.

Recently, Galvao and Wang (2015) proposed minimum distance estimator for the panel quantile regression that addresses the issue with increased computational costs of the Koenker (2004) estimator and more importantly avoids altogether the need to search for an optimal shrinkage. Their estimator for the slopes is defined as a weighted average of quantile regression estimators implemented over the individual observations with weights given by the inverses of individual covariance matrices. In principle, the estimator is carried out by estimating separate quantile regression for each individual (i.e. using its time series) and then averaging these estimates by using the inverse of estimated individual covariance matrices as weights. The first step involves solving a set of quantile regression problems, but these are much simpler and easier to estimate than the composite quantile regression problem being solved by the Koenker (2004) estimator. Then, the weighting scheme replaces the search for optimal shrinkage. Since the estimation of the covariance matrices needed to implement this weighting scheme is much simpler and easier this brings considerable computational savings. The minimum distance estimator of Galvao and Wang (2015) is the estimator we adopt in this paper.

### *3.3. Endogeneity*

Finally, we explicitly tackle the problem of potential endogeneity of the human capital factor variable.

There are three general approaches for dealing with endogeneity in semi- and nonparametric models. Using the terminology of Blundell and Powell (2003), these are the ‘fitted values’, the ‘instrumental variables’ and the ‘control function’ approaches. In quantile regression models, the ‘fitted values’ approach is exemplified by the two-stage quantile regression estimator of Kim and Muller (2004). The ‘instrumental variables’ approach has several different implementations such as the instrumental variables quantile regression method of Chernozhukov and Hansen (2006) or the moment based estimators of Kostov (2013). For the specifics of the control function approach to quantile regression see Lee (2004).

Although more generally applicable, the control function approach of Lee (2007) relies upon assumptions that are more difficult to test. The two-stage approach of Kim and Muller (2004) on the other hand imposes rather strong identification restrictions and is of limited use in potentially dependent data contexts. For these reason we opt for the

instrumental variable approach. The estimator of Chernozhukov and Hansen (2005, 2006) is such an instrumental variables approach, the assumptions of which are more easily aligned to the assumptions implied by linear models. Since, as we discuss below, there is some doubt about the strength of the instruments in growth regressions in general and in the particular formulation employed in this paper, it is important to be able to circumvent the problem of potentially weak instruments. Unfortunately, to the best of our knowledge, there is no formal test applicable to our quantile regression model. However the logic of the  $F$ -test rule of thumb routinely used in linear models would be applicable to the instrumental variables approach.

Second, the instrumental quantile regression approach has been used to extend the Koenker (2004) estimator to panel data models with endogeneity (see Galvao, 2011 and Harding and Lamarche, 2009, 2014). Consequently, as noted by Galvao and Wang (2015), the ‘fixed effects’ estimator of Koenker (2004) can also be represented as a minimum distance estimator and the theoretical proofs and arguments in this previous work carry forward to our application.

Third, the instrumental quantile regression is computationally attractive in carrying out a grid search by implementing a set of simple quantile regressions, which is reminiscent to the computational attractiveness of the minimum distance estimator.

Finally, using the control variables approach of Lee (2007) would have resulted in an overall estimator with the lowest computational costs. However, since in general the asymptotic distribution of control variable estimators in non- and semi-parametric models depends on the estimation method used in the structural part of the model (i.e. in equation (2) in this case) and since the minimum distance estimator we use a relatively new one, no results on this asymptotic distribution are currently available.

In order to implement the instrumental variable approach, one needs valid and relevant instruments. Although we have explained the appropriateness of the chosen instruments, one needs to take into account the possibility of them being weak. The concept and definition of “weak instruments” was formalized by Staiger and Stock (1997) and this work was rather influential in probably unintentionally popularising an informal rule of thumb in considering the  $F$ -statistic on the excluded instruments in the first stage. We say ‘unintentionally’, since the commonly used ‘critical’ value of 10 for the  $F$ -statistic could be misleading. Stock and Yogo (2005) discuss these issues in considerable detail and provide

useful rules of thumb regarding the weakness of instruments. They base the strength of the instruments on a concentration parameter measure and link this to the  $F$ -statistic and provide sets of critical values dependent on the error tolerance of the researcher and the number of excluded instruments.

Inference under weak identification is studied in Moreira (2003) and results have been extended beyond the linear model (see Stock and Wright, 2000; Kleibergen, 2005; Guggenberger et al, 2012). A useful overview of the underlying issues can be found in Andrews and Cheng (2012).

For the endogenous quantile regression case, Jun (2008) extended the results of Kleibergen (2005) for weak identification robust inference in general smooth moment condition models to the model of Chernozhukov and Hansen (2006). In this paper, we follow the related approach of Chernozhukov and Hansen (2008), which provides robust inference based on inverting a Wald type statistic. Although the approach of Chernozhukov and Hansen (2008) can also be adapted to construct formal tests on weak identification *via* projection, estimating the covariance matrix directly using their proposal has several distinct advantages. First, in this particular application it is the inference about the model parameters is of primary interest, not the issue of strength of the instruments. Second, since we apply a minimum distance estimator in which the instrumental variables quantile regression estimator is plugged-in for every individual (i.e. country) we actually need these robust estimates in order to calculate the weight for obtaining the final estimator. In applying a robust to weak identification inference we do not need to worry about the strength of the instruments, since we implicitly account for that in the estimation procedure.

We can therefore briefly summarise the estimation approach we have adopted.

First, we use a confirmatory factor analysis model (in equation (1)) to construct a latent factor measure for human capital. Then, we take this human capital measure as given and estimate the endogenous quantile regression model specified in equations (2) and (3). This model is estimated by the minimum distance estimator of Galvao and Wang (2015). More specifically, we estimate separate quantile regression equations for each individual (i.e. country) using time series data and obtain the final panel estimator by weighting these individual estimates by the inverses of their covariance matrices.

In order to account for the potential endogeneity of the human capital measure, we use the robust inference approach of Chernozhukov and Hansen (2008) in estimating the individual quantile regressions. This allows us to correct the individual estimates for endogeneity and more importantly to obtain robust to weak identification estimates for the country-specific covariance matrices. Hence, we use the covariance matrices implied by the robust estimation approach of Chernozhukov and Hansen (2008), instead of the asymptotic covariance matrices to construct the weights used to combine them in order to derive the final panel quantile regression estimator. Therefore, the above estimation approach allows us to carry the robustness to weak identification onto the final estimator (since the optimal weights are based on robust covariance estimation).

#### **4. Data**

We base our analysis on the well-known Mankiw et al. (1992) specification to evaluate the impact of saving, population growth, human capital on the growth rate of per capita income. Our dependent variable ( $y$  in equation (2)) is a 5-year average of growth rate of per capita GDP. Following Mankiw et al. (1992), the explanatory variables should be initial per capita GDP and proxies of average saving rate, population growth ( $X$  in equation (2)) and average human capital ( $w$  in equation (2)). These variables are drawn out of two databases.

On the one hand, the data for per-capita GDP, saving, population growth are extracted from the Heston et al. (2012) Penn World Table (PWT version 8.0), which contain information on real income, investment and population for a large number of countries. With respect to the control variables, we proxy population growth as the average growth of the working age population (15 to 64) on a basis of 5-year interval. The number of workers needed for the computation of this variable has been obtained as follows:  $RGDPCH * POP / RGDPW$ , where  $RGDPCH$  is real GDP per capita computed by the chain method,  $RGDPW$  is real-chain GDP per worker and  $POP$  is the total population. The savings rate is measured as the average share of gross investment in GDP for each five-year interval.

On the other hand, education related variables are extracted from The CANA database. This database contains 6 measures for education and human capital: educational

variables and three variables on gross enrolment ratios. The main educational variable 'Mean years of schooling' together with 'Public Expenditure on Education', and 'Primary teacher-pupil-ratio' are used to construct a confirmatory factor analysis human capital factor. 'Mean years of schooling' is the most widely used proxy for human capital. The other two variables are measures routinely used as human capital proxies in growth regressions. Moreover, they are both clearly related to the quality of the existing stock of human capital. 'Mean years of schooling' is related to the labour earnings approach to measuring human capital, while 'Public Expenditure on Education' is derived from the investment approach, and in view of the duality of these approaches, these two variables would not be normally used together. The conceptual approach we have presented earlier is however fully consistent with combining them since the latent human capital interpretation can effectively use information embodied in any proxies that measure this intrinsically unobservable concept.

The three gross enrolment ratios - primary, secondary and tertiary - are considered as instruments to account for the potential endogeneity of this factor variable. Because enrolment rates are aggregate determinants of education achievement they are appropriate as instruments. Note that educational enrolment is a preliminary step in building productive human capital, which will only materialise once the corresponding educational level is complete. Therefore, conceptually, education enrolment precedes the establishment of productive human capital capacity. This means that it is expected to be exogenous, but, on other hand, it will also be correlated with human capital. Because of the more roundabout way in which educational enrolment affects human capital, one may suspect weak instrumentation, something that we test and address.

Note that the assignment of 'Primary teacher-pupil-ratio' as a variable in the confirmatory factor analysis stage (i.e. in equation (1)) as a measure of human capital quality) instead of an instrument is slightly problematic. We view this as a measure of the quality of primary education and hence as a quality measure it should contribute to the quality of human capital directly, rather than in the more roundabout ways the instruments we use here are supposed to do. Although such a quality measure is bound to be more important for countries with lower stock of human capital (i.e. poorer nations) it can be reasonably assumed to contribute to the quality of human capital in more general terms. Assuming more or less coherent educational systems would mean that the



quality of primary education would in general be correlated with the quality of the other levels of education hence feeding into the quality of human capital. In addition to providing justification for considering this variable in the confirmatory factor analysis stage, the above discussion casts serious doubts about its appropriateness as an instrument.

In sum, the data used in this study consists of 5-year averages for the per-capita GDP, saving, population growth and human capital measures. These averages are calculated for each year over the previous five-year period. In this way, we obtain a dataset in which the variation in the data (in particular growth rates) is reduced by averaging, but nevertheless we have annual observations for each variable. In the empirical specification, we use the lagged value for the GDP per capita as initial income measure, which in this case is the value for the previous year. Owing to the way the latter is calculated, this is in fact the average value for the 5 year period preceding the current year. In other words, while the current value is calculated as an average for the last 5 years, including the current year, the lagged value is essentially calculated over the previous 5 years excluding the current year. Such aggregation is a standard practice in panel data growth studies and have been shown to improve significance of the result in accordance with theoretical predictions (Krueger and Lindahl, 2001).

These variables are constructed for an unbalanced sample of 124 countries over the period 1980-2010. Hence we have  $t = 1 \dots 29$  and  $N = 124$ . The countries in the main dataset and the years of availability are displayed in Table 1.

**Table 1. Countries in the dataset and years of availability**

Country	Years of availability	Country	Years of availability	Country	Years of availability
Albania	1980-2008	Greece	1980-2008	Oman	1980-2008
Angola	1980-2008	Guatemala	1980-2008	Pakistan	1980-2008
Argentina	1980-2008	Guinea	1985-2008	Panama	1980-2008
Armenia	1995-2008	Honduras	1980-2008	Paraguay	1980-2008
Australia	1980-2008	Hungary	1980-2008	Peru	1980-2008
Austria	1980-2008	Iceland	1980-2008	Philippines	1980-2008
Azerbaijan	1995-2008	India	1980-2008	Poland	1980-2008
Bahrain	1980-2008	Indonesia	1980-2008	Portugal	1980-2008
Bangladesh	1980-2008	Iran	1980-2008	Qatar	1980-2008
Belgium	1980-2008	Ireland	1980-2008	Romania	1980-2008
Benin	1985-2008	Israel	1980-2008	Russia	1995-2008
Bolivia	1980-2008	Italy	1980-2008	Rwanda	1985-2008
Botswana	1985-2008	Jamaica	1980-2008	Saudi Arabia	1980-2008
Brazil	1980-2008	Japan	1980-2008	Senegal	1980-2008
Bulgaria	1980-2008	Jordan	1980-2008	Sierra Leone	1985-2008
Burkina Faso	1980-2008	Kazakhstan	1995-2008	Singapore	1980-2008
Burundi	1985-2008	Kenya	1980-2008	Slovenia	1995-2008
Cambodia	1980-2008	Kuwait	1980-2008	South Africa	1980-2008
Cameroon	1980-2008	Kyrgyzstan	1995-2008	Spain	1980-2008
Canada	1980-2008	Latvia	1995-2008	Sri Lanka	1980-2008
Chad	1985-2008	Lebanon	1985-2008	Sudan	1980-2008
Chile	1980-2008	Lesotho	1985-2008	Swaziland	1985-2008
China	1980-2008	Liberia	1985-2008	Sweden	1980-2008
Colombia	1980-2008	Lithuania	1995-2008	Switzerland	1980-2008
Costa Rica	1980-2008	Madagascar	1980-2008	Tajikistan	1980-2008
Cote d'Ivoire	1980-2008	Malawi	1980-2008	Tanzania	1980-2008
Croatia	1995-2008	Malaysia	1980-2008	Thailand	1980-2008
Czech Republic	1995-2008	Mali	1980-2008	Togo	1985-2008
Denmark	1980-2008	Mauritania	1985-2008	Tunisia	1980-2008
Dominican Republic	1980-2008	Mauritius	1985-2008	Turkey	1980-2008
Ecuador	1980-2008	Mexico	1980-2008	Uganda	1980-2008
Egypt	1980-2008	Moldova	1995-2008	Ukraine	1995-2008
El Salvador	1980-2008	Mongolia	1985-2008	United Kingdom	1980-2008
Estonia	1995-2008	Morocco	1980-2008	Uruguay	1980-2008
Ethiopia	1980-2008	Mozambique	1980-2008	Uzbekistan	1995-2008
Fiji	1985-2008	Namibia	1985-2008	Venezuela	1980-2008
Finland	1980-2008	Nepal	1985-2008	Vietnam	1980-2008
France	1980-2008	Netherlands	1980-2008	Yemen	1994-2008
Gabon	1985-2008	New Zealand	1980-2008	Zambia	1980-2008
Georgia	1995-2008	Niger	1980-2008	Zimbabwe	1980-2008
Germany	1980-2008	Nigeria	1980-2008		
Ghana	1980-2008	Norway	1980-2008		

In addition to the main sample, in order to investigate efficiency gains, we also construct a small sub-sample. The small subsample is obtained by sub-setting the main sample at 5 years intervals. That is instead of taking each annual value we take the 1980, 1985, etc. values only (with 2008 instead of 2010, to gain another observation since the CANA database ends in 2008). This results in considerably smaller subset of the data with  $t = 1 \dots 7$ . The construction of a small sample in the time dimension is typical of some previous cross-country studies (as in e.g. Burnside and Dollar, 2000; Dalgaard et al., 2004) and since it is a subset of the main sample we use here, it is clear that a model based on such small sample will suffer efficiency losses. It would therefore be informative to be able to shed some light on the nature of such efficiency losses.

## 5. Results

Before proceeding to the main results we briefly discuss the construction and the interpretation of the latent human capital measure. In essence we conflate several human capital measures into a single factor. Although such a practice could at first sight appear to contradict the discussion of the multidimensional nature of human capital itself, economic theory posits strong implication on the effects of human capital that are easier tested by combining such different measures into a single one. Furthermore, this is consistent with the previous growth literature that uses a single proxy for that purpose. The idea of extracting a single latent factor means that we are extracting the common source of variation in our human capital proxies, which can be hypothesised to be common driving factor for these measures. This interpretation of human capital as a latent factor sits naturally within the latent variables models, the simplest one of which is the confirmatory factor analysis (CFA). However, the purpose of the present approach is slightly different to that of a conventional CFA. We simply want to extract the common variance present in the measurement variables. In a way this is no different from deriving the most important factor in an exploratory factor analysis. In this particular case we only have a single latent variable with three human capital measures. This means that the latent factor model is just identified, which precludes the calculation of validity measures (comparative fit indices), since these rely on a comparison with a simpler null model.

The estimated human capital factor is presented in Table 2. It is standardised so that the contribution of the most widely used proxy, namely ‘Mean Years of Schooling’ is set to 1. Interestingly the contribution of ‘Public expenditure on education’ is considerably lower. To properly account for the contributions of these different measure we need to weigh them by their variances. Doing so and scaling these to add up to 1, we end up with a measure of the relative importance (i.e. relative contribution) of each measure, presented in the last column of Table 2. One could get similar outcome if all measures were standardised (e.g. scaled to a common variance) prior to the analysis. The relative importance shows that although ‘Public expenditure on education’ appears to contribute relatively little to the estimated human capital factor, perhaps surprisingly the contribution of ‘Primary teacher-pupil-ratio’ is comparable (and even slightly higher) than this of the main proxy ‘Mean Years of Schooling’. The high contribution of ‘Primary teacher-pupil-ratio’ is an interesting finding, which to some extent can be explained by the composition of the data sample with high proportion of developing countries where one can hypothesis such a variable would have large effect of human capital. This finding is nevertheless not so unexpected since the arguments in favour of these two measures have been extensively discussed in the empirical literature.

**Table 2. Latent factor estimation results**

	Estimate	SE	P-Value	Relative importance
<b>Contributions to HC</b>				
Mean years of schooling	1.00			0.43
Public Expenditure on Education	0.35	0.02	0.00	0.11
Primary teacher-pupil-ratio	6.68	0.36	0.00	0.46
<b>Variances:</b>				
Mean years of schooling	3.18	0.22	0.00	
Public Expenditure on Education	4.40	0.11	0.00	
Primary teacher-pupil-ratio	19.93	8.95	0.03	
Human Capital	3.85	0.25	0.00	

With respect to our instrumentation strategy, we first display in Table 3 some standard diagnostics. When using all enrolment ratios as instruments (first part of Table 3), the Wu-Hausman test confirms the endogeneity of the human capital variables. Furthermore, the null hypothesis of weak instrument is strongly rejected. However, the Sargan test for over-identification is marginally significant, indicating that the moment orthogonality

conditions may be violated. Despite the fact that the Sargan test may lack power (Parente and Santos Silva, 2012), we have systematically looked for all the subsets of these instruments that do not imply a rejection of the null of moment orthogonality. It appears that when dropping the primary enrolment rates for the set of instruments, the Sargan test does not reject moment orthogonality. The remaining two enrolment rates still have great explanatory power, as shown by the second part of Table 3.

**Table 3. Diagnostics for instruments**

	Test statistic	df	<i>p</i> -value
<b>All enrolment ratios:</b>			
Weak instruments	121.03	(3,676)	0.00
Wu-Hausman	9.04	(1,677)	0.00
Sargan	6.66	1	0.04
Encompassing J test			
M1 vs. ME	26.74	572	0.00
M2 vs. ME	2.04	572	0.15
<b>Two enrolment ratios</b>			
Weak instruments	175.14	(2,676)	0.00
Wu-Hausman	11.91	(1,677)	0.00
Sargan	0.21	2	0.64
Encompassing J test			
M1 vs. ME	34.84	572	0.00
M2 vs. ME	1.84	572	0.18
Wald test of joint significance of instruments	16.60		0.00

We also test the endogeneity of the human capital variable via an encompassing J test. In simple terms we want to compare the standard model assuming that variables are exogenous (M1) vs a model in which the endogenous variable is replaced by its linear projection onto the instruments (M2 which is the IV model). In order to compare the predictive power of these two models an encompassing model (ME) is constructed containing both the endogenous variable and its linear projection. Then the tests statistics are just Wald tests for models M1 and M2 vs ME. If the human capital is exogenous then the instruments should have no additional explanatory power and hence the human capital projection onto the vector space of instruments should be insignificant. Also the model M2 should be equivalent to M1. In both cases (two or three enrolment

ratios), the test results demonstrate that the human capital variable is endogenous and that the instruments have explanatory power. Finally we additionally test the joint explanatory power of the final set of instruments i.e. without the primary enrolment ratios) via a Wald test which confirms their validity.

We now proceed to the description of the main estimation results. We use a grid of 91 quantiles from 0.05 to 0.95 at 0.01 increments to approximate the conditional growth distribution. The estimated coefficients for each of the growth determinants across these quantiles are plotted on Figure 1 and provide full characterisation of the conditional growth process. In particular, the upper conditional quantiles represent the countries which, given their endowments grow faster than the other comparable countries while the lower conditional quantiles represent the countries, which grow slower than they should compared to similar countries.

Additionally, Table 4 presents the estimated coefficients and their standard errors for a subset of quantile models (at each 10th percentile), together with two mean models, namely the standard instrumental variable model labelled as IV1 and IV estimation with an additional non-linear instrument based on savings, following Lewbel (2012), labelled as IV2. The latter is the optimal non-linear instrument choice decided upon by testing all possible combinations involving the three exogenous variables employed in the model specification.

**Table 4. Comparison of mean and quantile estimation results**

	Human capital	SE	Initial income	SE	Population growth	SE	Savings	SE
IV1	0.0183	0.0030	-0.0297	0.0047	0.0027	0.0014	0.0027	0.0014
IV2	0.0186	0.0032	-0.0302	0.0054	0.0028	0.0014	0.0208	0.0050
$q = 0.1$	0.0391	0.0040	-0.0388	0.0046	0.0028	0.0012	0.0324	0.0062
$q = 0.2$	0.0338	0.0031	-0.0345	0.0036	0.0037	0.0009	0.0191	0.0038
$q = 0.3$	0.0273	0.0026	-0.0287	0.0031	0.0033	0.0008	0.0142	0.0029
$q = 0.4$	0.0240	0.0022	-0.0270	0.0028	0.0023	0.0008	0.0106	0.0025
$q = 0.5$	0.0228	0.0020	-0.0286	0.0026	0.0020	0.0007	0.0098	0.0022
$q = 0.6$	0.0221	0.0020	-0.0304	0.0027	0.0012	0.0007	0.0116	0.0021
$q = 0.7$	0.0214	0.0020	-0.0318	0.0028	0.0009	0.0006	0.0148	0.0019
$q = 0.8$	0.0210	0.0021	-0.0347	0.0031	-0.0009	0.0007	0.0164	0.0021
$q = 0.9$	0.0216	0.0026	-0.0415	0.0040	-0.0047	0.0009	0.0200	0.0032

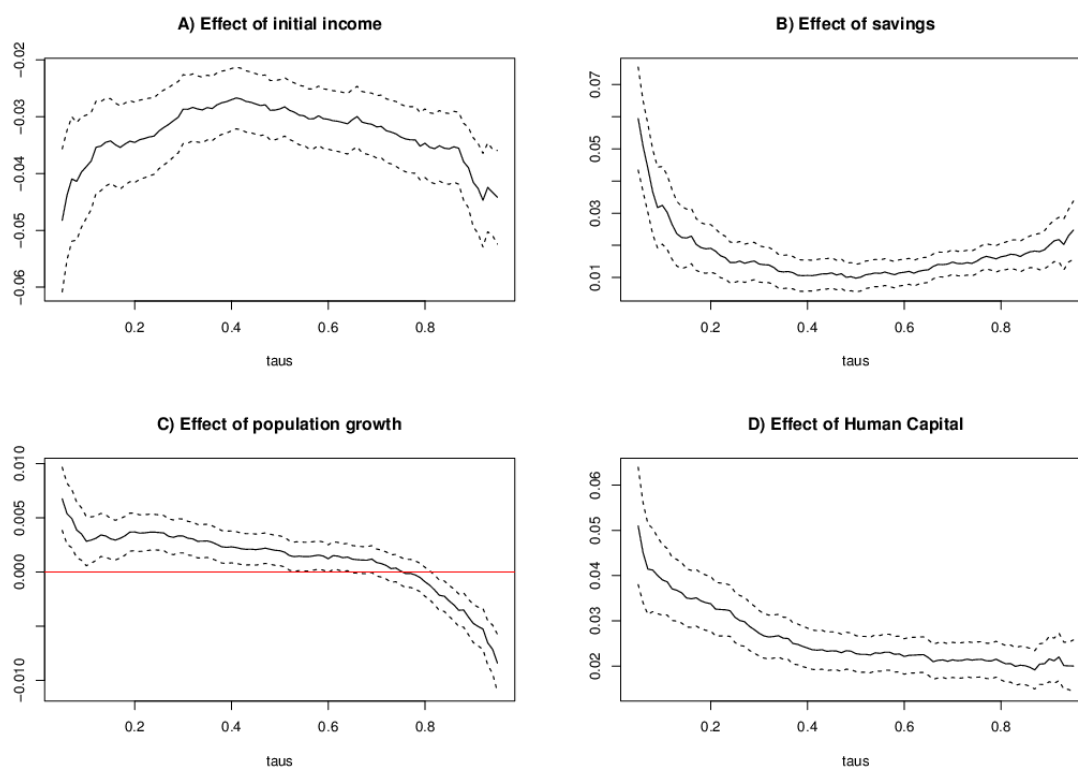
In general terms, after taking into account the variation of estimated effect across different quantiles, the results seem comparable to the mean models with two notable exceptions. First, the human capital effects are larger in the quantile model, for all quantiles, relative to those in the mean models. Second the effect of saving in the standard mean model (IV1) is considerably lower than the same effect in both IV2 and the quantile models. Furthermore, unlike the other models in IV1, this coefficient is only marginally significant (exact  $p$ -value of 0.059). By providing robust instrumentation, both IV2 and the quantile models obtain comparable in magnitude and significance effects for the savings variable.

For ease of interpretation in the rest of the paper the estimated effects are presented in graphical form<sup>2</sup>. The first point of interests is the effect of the lagged income (see Figure 1A) since it provides direct link to the conditional convergence hypothesis. Because the upper quantiles signify conditionally faster growing countries, we expect the coefficient of lagged income to be negative and increasing in magnitude with quantiles. Indeed, overall, we obtain a significantly negative coefficient, which provides supporting evidence for conditional convergence. In general, such a finding is consistent with Cunningham (2003), Barreto and Hughes (2004), Canarella and Pollard (2004) and Kostov and Le Gallo (2015). Furthermore, in contrast to those previous studies (which show insignificant coefficients at some quantiles), we obtain significantly negative coefficients across the whole conditional distribution. It can therefore be claimed that we find stronger support for the conditional convergence hypothesis. However, since we have implemented several additional elements not present in these previous studies (taking account of individual effects in our panel setting, latent factor formulation for the human capital variable and taking account of the endogeneity issue for the latter), it is difficult to disentangle the differences alongside these distinct methodological innovations. Nevertheless, it is clear that the present study presents stronger support for conditional convergence.

---

<sup>2</sup> Reproducing them in tabular format as in Table 4 would be impractical since it would involve 91 rows for each set of estimations. Full numerical results are available upon request from the authors.

**Figure 1. Estimation results**



It terms of empirical modelling the specifications of Kostov and Le Gallo (2015) are closest to the one employed in the presents study and therefore easier to compare. They have also used panel quantile regression, but applied a different measure for human capital (from Barro and Lee, 2010) and did not take into account its potential endogeneity. Hence any differences between our results and those of Kostov and Le Gallo (2015) can be attributed to the two main methodological innovations, namely the human capital measure and accounting for its endogeneity. In contrast to the results of Kostov and Le Gallo (2015) the lagged income effect (see Figure 1A) is not only significant over the whole conditional distribution, but also considerably larger in magnitude.

The effect of savings (Figure 1B) is positive (as expected) and is greater in the left tail, meaning that in (conditionally) slower growing economies savings contribute more to growth. Furthermore, one can also draw a tentative (due to the width of the confidence intervals) conclusion about possibly stronger effect of savings in the right tail. This finding is at odds with the results of Barreto and Hughes (2004) who reported that the effect of investment share increases with quantiles. Interestingly, our results are not entirely inconsistent with Canarella and Pollard (2004) who only found significant



coefficients for lower and higher quantiles. Because the effect of saving is higher in the lowest quantiles, our findings are consistent with the conjecture that conditionally slower growing economies may suffer from insufficient capital accumulation hence leading to larger marginal effect of savings. Both in terms of magnitude and overall pattern our findings about the effect of savings are consistent with those of Kostov and Le Gallo (2015).

Growth in working population has generally positive effect on growth (Figure 1C), but this effect reduces with quantile and only exists in the lower half of the growth distribution. Population growth is insignificant between the 0.5<sup>th</sup> and 0.8<sup>th</sup> quantiles and negative beyond that. This could be interpreted as an evidence that increasing workforce contributes to economic growth but up to a certain point. After that, the marginal effect of workforce is not a significant determinant of growth. For conditionally slower growing economies this effect is larger while it reduces and totally disappears for faster growing ones, which do not base their growth on labour. Indeed, for conditionally faster growing economies it is the quality (see next result) rather than the quantity of the workforce that accelerates growth. Once again the results with regard to population growth are very similar to those of Kostov and Le Gallo (2015).

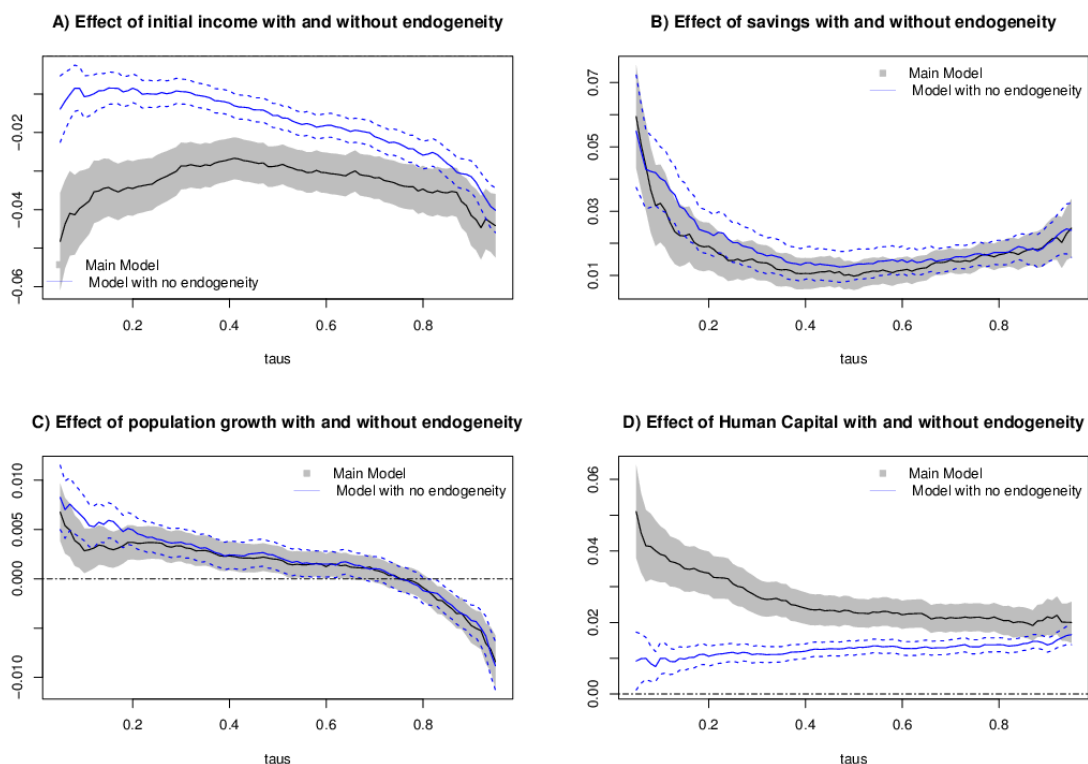
Human capital, measured with the latent factor measure, has a positive and significant effect for all quantiles of the conditional growth distribution (Figure 1D). This effect appears to be greater in the left tail, implying that relatively slower growing economies can benefit to larger extent from a better quality of human capital. Comparisons with previous studies can only be tentative since such studies have employed a particular proxy for human capital and their results can be expected to be heavily influenced by their particular choice of such proxy and the fact that the proxy does not capture quality. Barreto and Hughes (2004) found that the importance of secondary school attainment reaches a peak between the 30<sup>th</sup> percentile and the median and then drops for higher percentiles. Canarella and Pollard (2004) on the other hand found that the estimates for human capital (proxied by the average fraction of working age population in secondary school between 1960 and 2000) increase with quantiles and are not significant for lower quantiles. Our results are closer to the theoretical expectations in that the coefficients are significantly positive across the whole distribution. Kostov and Le Gallo (2015) only find a significant effect for human capital under an ‘endogenous spatial’ specification. In a

nutshell they specify a growth spillover process following a spatial econometrics model in which they specify a spatial lag of the growth variable. The distances between countries (and hence their growth interdependence) are calculated based on the so called market potential (Redding and Venables, 2004; Head and Mayer (2004). The latter is based on the level of factor incomes, weighted by bilateral trade costs. Since the only factor considered in their calculations is labour i.e. factor incomes are the wages) the resulting spatial models specifies growth spillovers with regard to labour earning capacity, which on its own can be considered as a measure of human capital. Therefore we can draw a direct parallel to their results since, in a way, they provide (via a complicated econometric specification) a better measure for human capital. Yet the human capital effects we estimate are even stronger.

We now proceed to a set of comparisons of our results relative to some alternative model formulations in order to ascertain the importance of different modelling assumptions for the results. First, we look at the effect of correcting for the potential endogeneity of the human capital measure. Figure 2 contrasts the main results against a model in which the endogeneity issue is ignored. In any other respect, this alternative model is identical to the main one. The differences in the estimates and confidence intervals for savings and population growth (Figure 2B and 2C) are relatively small. However, taking into account endogeneity affects the estimates for human capital (as it is to be expected, see Figure 2D) and the convergence rates (Figure 2A). Ignoring endogeneity appears to underestimate both convergence rates and the effect of human capital. Note that both these are movements ‘away’ from the theoretical expectations. Also, as expected, the confidence intervals from the endogenous model are wider and again this seems to only apply to the effects of HC and initial income, although in this case it does not lead to any qualitative consequences (as it could have been the case if some of the effects were insignificant since then wider confidence intervals may have dragged more quantiles into the insignificance region). There is however an important qualitative difference in these results in what refers to the convergence rates. The effects of the initial income for the exogenous model formulation appear to be downward sloping. This means that conditionally higher growing economics are converging faster, which would be the prediction from the ‘pure’ convergence hypothesis and would be inconsistent with the notion of e.g. convergence clubs, because it would rule out multiple steady states. This is the general pattern of the estimates of Kostov and Le Gallo (2015). The main model

results however do not exhibit this shape. Therefore, although ignoring endogeneity appears to be closer to the original convergence hypothesis, our results appear more realistic in that they do not rule out the possibility of multiple points of convergence. Interestingly the exogenous model results for human capital are quite similar to those of the ‘endogenous spatial specification’ of Kostov and Le Gallo (2015).

**Figure2. Effects of ignoring endogeneity**

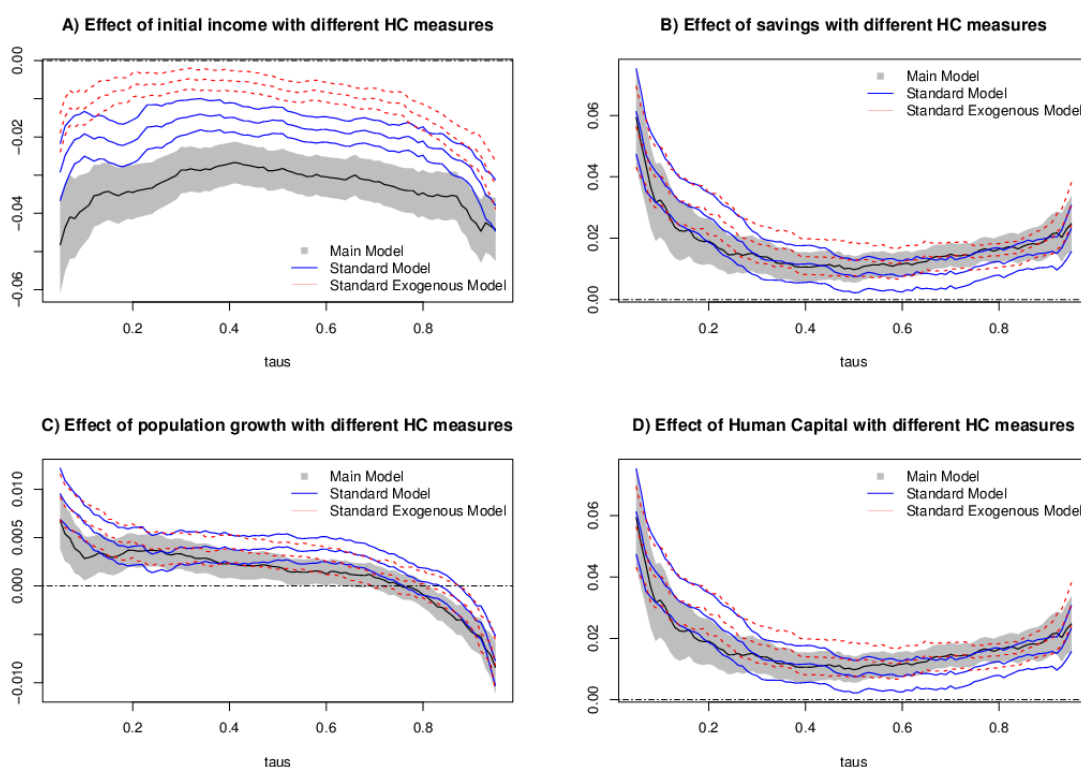


The next effect we would like to investigate is how the introduction of the latent factor measure for human capital affects the results. Figure 3 compares the main model with two alternative models. These are both based on ‘Mean years of schooling’ as a human capital measure. Since this is probably the most popular human capital proxy in empirical studies, it is informative to compare it with our results. We have undertaken similar comparisons with the other two proxies (used to construct the latent factor measure) and the conclusions are qualitatively similar.<sup>3</sup> We employ two alternative models in this comparison. The first one is a ‘Standard model’, which is identical to the main one except the human capital measure, e.g. it uses the same set of instruments and estimation

<sup>3</sup> They are available upon request from the authors.

method to correct for endogeneity. The other model is a ‘Standard Exogenous model’ which in addition to using the alternative human capital measure also ignores the endogeneity issue (hence compounding the effects of alternative HC measure and endogeneity).

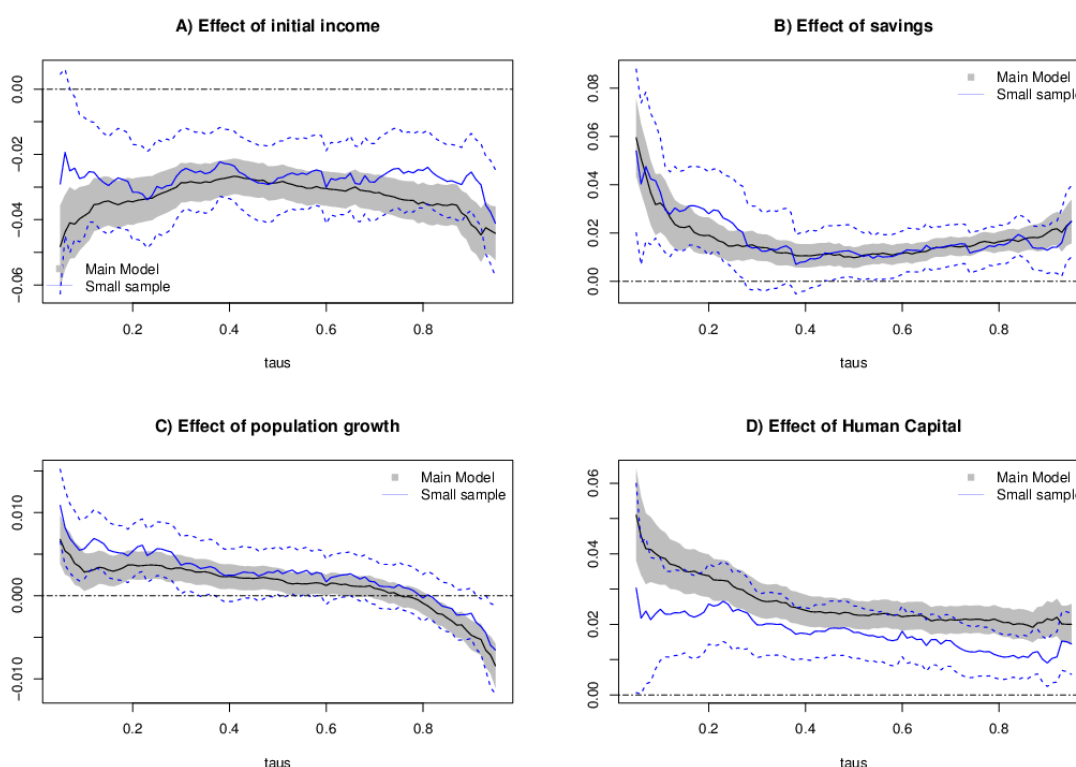
**Figure 3. Results under different HC measures**



Using the ‘standard’ measure of HC leads to considerably lower estimated convergence rates and the difference is larger when the endogeneity is not accounted for. The effects of the different HC measure are broadly similar, for the other variables but with some important qualitative difference. In particular in the standard model (with endogeneity) the effects of savings and human capital edge closer to insignificance (although they still remain significant) in the middle of the conditional distribution. This means that qualitative difference may appear with different sample (where confidence interval may become wider). In the case of population growth the alternative HC measure models are more supportive of its positive effect (in that the effect becomes insignificant only at around the 0.8<sup>th</sup> quantile and does not turn negative until the 0.9<sup>th</sup>).

Finally, Figure 4 compares the main model with the small sample (in the time dimension) version. Since, as discussed in the data section, the small sample is simply a subset of the data used in the main model, it is clear that it will lead to some inefficiency in the estimation. In particular, since the time dimension of the small sample is only 7 observations, such inefficiencies could be expected to be considerable in the case of the estimator we employ in this paper.

**Figure 4. Main results vs small sample**



First of all, one may notice that although the estimated effects with exception of those for the human capital itself are broadly similar between the two samples, the confidence intervals for the small sample are drastically wider. These wider confidence intervals lead to a larger part of the conditional distribution for the effects of population growth being dragged into the insignificance region, and even some of the savings effects become insignificant. Interestingly, the human capital marginal effect in the small sample looks smaller than those in the main model. Hence, although in principle reducing the sample following averaging may not affect the point estimates too much, due to the drastic differences in the estimated confidence intervals, qualitative inference may lead to quite different conclusions. In particular the small sample equivalents of Figures 2 and 3

(available from the authors) reveal considerably more differences than those we considered in the paper.

## **6. Conclusion**

The aim of this paper is to reconsider the way human capital has been incorporated in cross-country growth studies. Reflecting on the critiques of alternative proxy measures for human capital employed in previous research, we take a radically different approach. Following the approach set out in Dagum and Slottje (2000) at a microeconomic level, we employ a confirmatory factor analysis to construct a new measure for human capital by combining several widely used proxies and use the measure constructed in this way in a panel quantile cross-country growth regression to investigate its impact on aggregate growth. We also take into consideration the possible endogeneity of this measure and the weak identification issues. Furthermore, instead of focusing on the average form of the relationship, we analyse the whole conditional growth distribution and account for country unobserved heterogeneity through fixed effects.

Our results are supportive of the conditional convergence hypothesis and the signs of the explanatory variables are as expected. Furthermore, our findings tend to lend more support to conditional convergence, for all quantiles of the conditional growth distribution, compared to previous applications of similar methodologies. More importantly, our results show significant effects of human capital, suggesting that the imperfect measures used in previous studies may to a large extent be responsible for the counterintuitive and contradictory findings for its effect.

## References

- Abrevaya J., Dahl C.M. (2008) The effects of birth inputs on birthweight: evidence from quantile estimation on panel data, *Journal of Business and Economic Statistics*, 26, 379–397.
- Ando T., Tsay R.S. (2011) Quantile regression models with factor-augmented predictors and information criterion, *The Econometrics Journal*, 1–24.
- Andrews D., Cheng X. (2012) Estimation and inference with weak, semi-strong, and strong identification, *Econometrica*, 80, 2153–2211.
- Barreto R.A., Hughes A.W. (2004) Under performers and over achievers: a quantile regression analysis of growth, *The Economic Record*, 80(248), 17–35.
- Barro R.J., Lee J. (2010) A new data Set of educational attainment in the world, 1950–2010, NBER Working Paper 15902.
- Brown T.A. (2015) *Confirmatory Factor Analysis for Applied Research*, Guilford Publications.
- Brumm H.J. (1997) Military spending, government disarray, and economic growth: a cross-country empirical analysis, *Journal of Macroeconomics*, 19, 827–838.
- Burgette L.F., Reiter J.P. (2012) Modeling adverse birth outcomes via confirmatory factor quantile regression, *Biometrics*, 68, 92–100.
- Burnside C., Dollar D. (2000) Aid, policies, and growth, *American Economic Review*, 90(4), 847–868.
- Canarella G., Pollard S. (2004) Parameter heterogeneity in the neoclassical growth model: a quantile regression approach, *Journal of Economic Development*, 29(1), 1–31.
- Canay I.A. (2011) A simple approach to quantile regression for panel data, *Econometrics Journal*, 14, 368–386.
- Castellacci F., Natera J.M. (2011) A new panel dataset for cross-country analysis of national systems, growth and development (CANA), *Innovation and Development*, 1, 205–226.
- Chernozhukov V., Hansen C. (2005) An IV model of quantile treatment effects, *Econometrica*, 73, 245–261.
- Chernozhukov V., Hansen C. (2006) Instrumental quantile regression inference for structural and treatment effects models, *Journal of Econometrics*, 132, 491–525.
- Chernozhukov V., Hansen C. (2008) Instrumental variable quantile regression: A robust inference approach, *Journal of Econometrics*, 142(1), 379–398.

- Cohen D., Soto M. (2007) Growth and human capital: good data, good results, *Journal of Economic Growth*, 12, 51–76.
- Cracolici M.F., Cuffaro M., Nijkamp P. (2010) The measurement of economic, social and environmental performance of countries: a novel approach, *Social and Indicator Research*, 95, 339–356.
- Cragg J.G., Donald S.G. (1993) Testing identifiability and specification in instrumental variable models, *Econometric Theory*, 9, 222–240.
- Cunningham B.M. (2003) The distributional heterogeneity of growth effects: some evidence, *The Manchester School*, 71, 417–447.
- Dagum C., Slottje D.J. (2000) A new method to estimate the level and distribution of the household human capital with application, *Structural Change and Economic Dynamics* 11, 67–94.
- Dagum C. (1994) Human Capital, income and wealth distribution models and their application to the U.S.A., *Proceedings of the Business and Economic Statistic Section of the American Statistical Association*, 154th Meeting, 253–258.
- Dagum C., Vittadini G., Lovaglio P.G. (2007) Formative indicators and effects of a causal model for household human capital with applications, *Econometric Review* 26(5), 579–596.
- Dalgaard C.-J., Hansen H., Tarp F. (2004) On the empirics of foreign aid and growth, *Economic Journal*, 114, F191–F216.
- de la Fuente A., Domenech R. (2006) Human capital in growth regressions: how much difference does data quality make? *Journal of the European Economic Association*, 4(1), 1–36.
- Delgado M.S., Henderson D.J., Parmeter C.F. (2014) Does education matter for economic growth? *Oxford Bulletin of Economics and Statistics*, 76(3), 334–359.
- Dolton P., Marcenaro-Gutierrez O.D. (2011) If you pay peanuts do you get monkeys? A cross-country analysis of teacher pay and pupil performance, *Economic Policy*, 26(65) 5–55.
- Dufrenot G., Mignon V., Tsangarides C. (2010) The trade-growth nexus in the developing countries: a quantile regression approach, *Review of World Economics*, 146, 731–761.
- Durlauf S.N., Johnson P.A., Temple, J.R.W. (2009) The Econometrics of Convergence, in Terence C. Mills and Kerry Patterson (eds.) *Palgrave Handbook of Econometrics, Volume 2: Applied Econometrics*. Palgrave Macmillan, June.



- Eisner R. (1985) The total incomes system of accounts, *Survey of Current Business*, 65, 24-48.
- Engel E. (1883) *Der Werth des Menschen*, Verlag von Leonhard Simion, Berlin.
- Földvári P., van Leeuwen B. (2006) An estimation of the human capital stock in Eastern and Central Europe, *Eastern European Economics*, 43(6), 53-65.
- Folloni G., Vittadini G. (2010) Human capital measurement: a survey, *Journal of Economic Surveys*, 24(2), 248-279.
- Foster N. (2008) The impact of trade liberalisation on economic growth: evidence from a quantile regression analysis, *Kyklos*, 61(4), 543-567.
- Galvao A.F. (2011) Quantile regression for dynamic panel data with fixed effects, *Journal of Econometrics*, 164(1), 142-157.
- Galvao A.F., Wang L. (2015) Efficient minimum distance estimator for quantile regression fixed effects panel data, *Journal of Multivariate Analysis*, 133, 1-26.
- Guggenberger P., Ramalho J., Smith R. (2012) GEL statistics under weak identification, *Journal of Econometrics*, 170, 331-349.
- Hanushek E.A. (2016) Will more higher education improve economic growth, *Oxford Review of Economic Policy*, 32(4), 538-552.
- Hanushek E.A., Piopiunik M., Wiederhold S. (2018) The value of smarter teachers. International evidence on teacher cognitive skills and student performance, *Journal of Human Resources*, forthcoming
- Harding M., Lamarche C. (2009) A quantile regression approach for estimating panel data models using instrumental variables, *Economics Letters*, 104(3), 133-135.
- Harding M., Lamarche C. (2014) Estimating and testing a quantile regression model with interactive effects, *Journal of Econometrics*, 178, 101-113.
- Harrington D. (2008) *Confirmatory Factor Analysis*, Oxford University Press, Oxford.
- Head K., Mayer T. (2004) Market Potential and the location of Japanese investment in the European Union, *The Review of Economics and Statistics*, 86(4), 959-972.
- Jun S.J. (2008) Weak identification robust tests in an instrumental quantile model, *Journal of Econometrics*, 144(1), 118-138.
- Kalaitzidakis P., Mamuneas T.P., Savvides A., Stengos T. (2001) Measures of human capital and nonlinearities in economic growth, *Journal of Economic Growth*, 6, 229-254.
- Kendrick J. (1976) *The Formation and Stocks of Total Capital*, Columbia University Press, New York.

- Kim, T. H. and C. Muller (2004). Two-stage quantile regression when the first stage is based on quantile regression, *Econometrics Journal*, 7(1), 218-231.
- Kleibergen F. (2005) Testing parameters in GMM without assuming they are identified, *Econometrica*, 73, 1103–1123.
- Koenker R. (2004) Quantile regression for longitudinal data, *Journal of Multivariate Analysis*, 91, 74-89.
- Kostov, P. (2013) Empirical likelihood estimation of the spatial quantile regression, *Journal of Geographical Systems*, 15(1), 51-69.
- Kostov, P., Le Gallo, J. (2015) Convergence: a story of quantiles and spillovers, *Kyklos*, 68(4), 552-576.
- Lamarche C. (2010) Robust penalized quantile regression estimation for panel data, *Journal of Econometrics*, 157(2), 396-408.
- Lee, S (2007) Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics*, 141 (2) 1131 – 1158.
- Le T. (2006) Estimating the monetary value of the stock of human capital for New Zealand. PhD Thesis, University of Canterbury.
- Le T., Gibson J., Oxley L. (2003) Cost- and income-based measures of human capital, *Journal of Economic Surveys*, 17(3), 272-307.
- Lee S. (2007) Endogeneity in quantile regression models: A control function approach, *Journal of Econometrics*, 141(2), 1131-1158.
- Lewbel, A. (2012). Using Heteroskedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models, *Journal of Business and Economic Statistics*, 30(1), 67-80.
- Lutz W., Goujon A., Samir K.C., Sanderson W. (2007) Reconstruction of populations by age, sex, and level of educational attainment for 120 countries for 1970–2000, *Vienna Yearbook of Population Research*, 5, 193–235.
- Mincer J. (1970) The distribution of labor incomes: a survey, *Journal of Economic Literature*, 8(1), 1–26.
- Mincer J. (1974) *Schooling, Experience, and Earnings*, New York: NBER Press.
- Moreira M. (2003) A conditional likelihood ratio test for structural models, *Econometrica*, 71, 1027–1048.

- Nehru V., Swanson E., Dubey A. (1995) A new database on human capital stock in developing and industrial countries: sources methodology and results, *Journal of Development Economics*, 46, 379–401.
- OECD (1998) *Human Capital Investment. An International Comparison*, Center for International Research and Innovation, Paris.
- Pritchett L. (2001) Where has all the education gone? *The World Bank Economic Review*, 15, 367-391.
- Ram R. (2008) Parametric variability in cross-country growth regressions: an application of quantile-regression methodology, *Economics Letters*, 99, 387-389.
- Redding, S., Venables A. (2004) Economic geography and international inequality, *Journal of International Economics*, 62(1), 53–82.
- Rosen A. (2009) Set identification via quantile restrictions in short panels CeMMAP working papers CWP26/09, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.
- Staiger D., Stock J. H. (1997) Instrumental variables regression with weak instruments, *Econometrica* 65(3), 557-586.
- Stock J.H., Wright J. (2000) GMM with weak identification, *Econometrica* 68, 1055–1096.
- Stock J.H., Yogo M. (2005) Testing for weak instruments in linear IV regression, in J.H. Stock and D.W.K. Andrews (eds), *Identification and Inference for Econometric Models: Essays in Honor of Thomas J. Rothenberg*, Cambridge: Cambridge University Press.
- Sunde U., Vischer T. (2015) Human capital and growth: specification matters, *Economica*, 82(326), 368-390.
- Temple J. (1999) The new growth evidence, *Journal of Economic Literature*, Vol. XXXVII, 112-156.
- United Nations, Department of Economic Affairs (1953) Concept and Definitions of Capital Formation, *Studies in Methods, series F*, No. 3.
- Vittadini G., Lovaglio, P.G. (2007) Evaluation of the Dagum–Slottje method to estimate household human capital, *Structural Change and Economic Dynamics*, 18(2), 270–278.
- Wössman L. (2003) Specifying human capital, *Journal of Economic Surveys*, 17(3), 239-270.
- Wössman L. (2005) Education production in Europe, *Economic Policy*, 20(43), 446-504.